

Using RDKit in Cheminformatics Visualizations

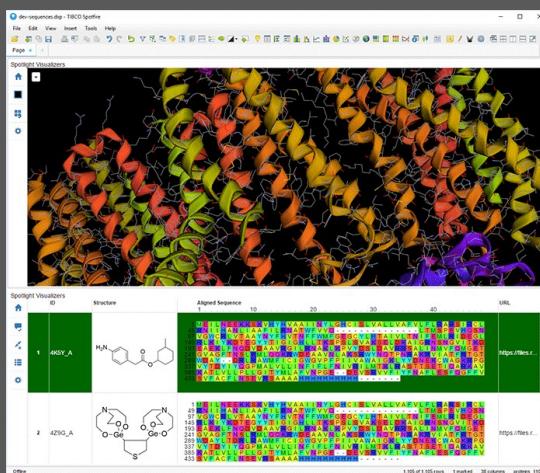
Gareth Jones

Glysade, LLC

2020 RDKit UGM

Lead Discovery ChemCharts

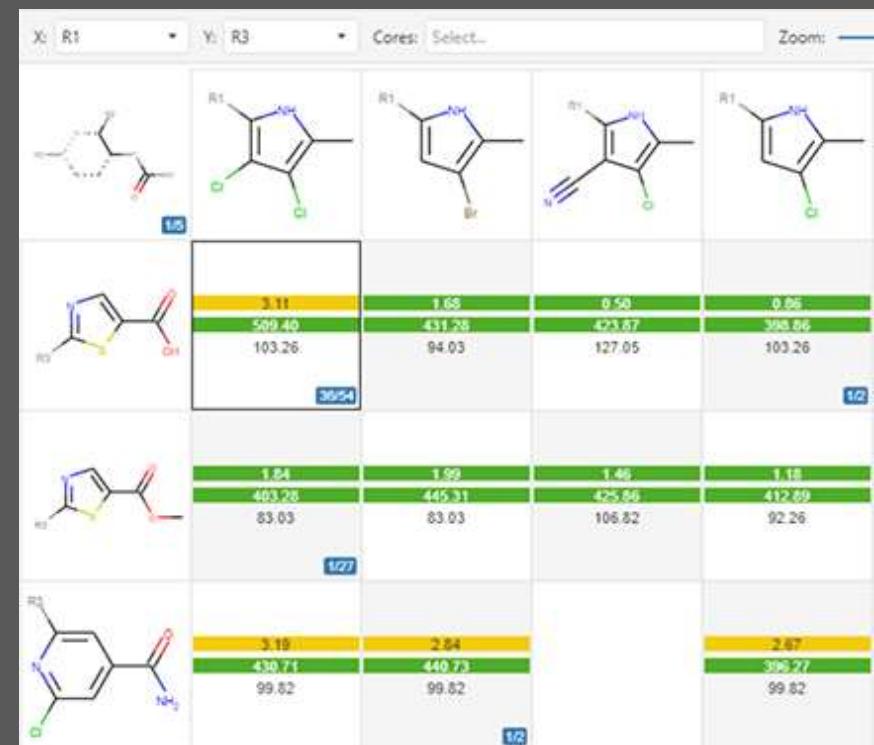
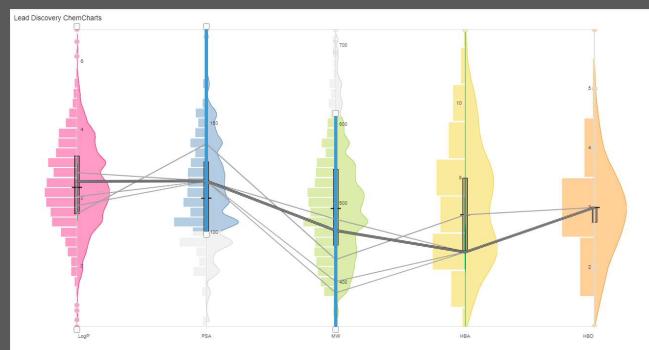
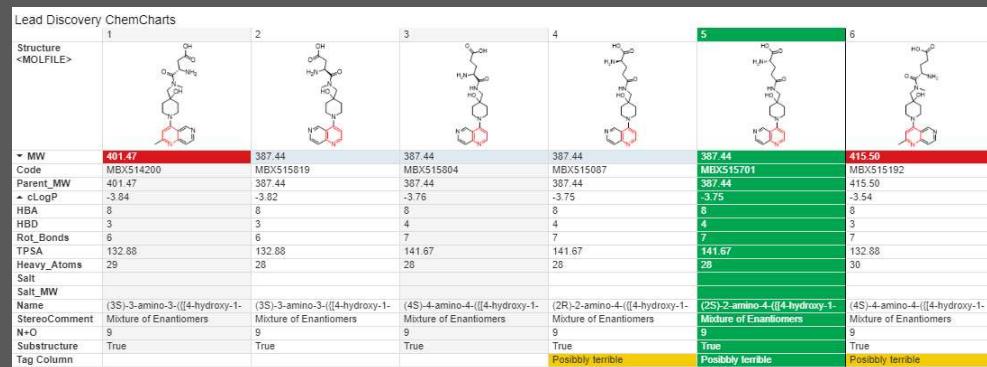
- Spotfire Plots for Chemistry and Biologics R&D
- Available through PerkinElmer



www.glysade.com/lead-discovery-premium

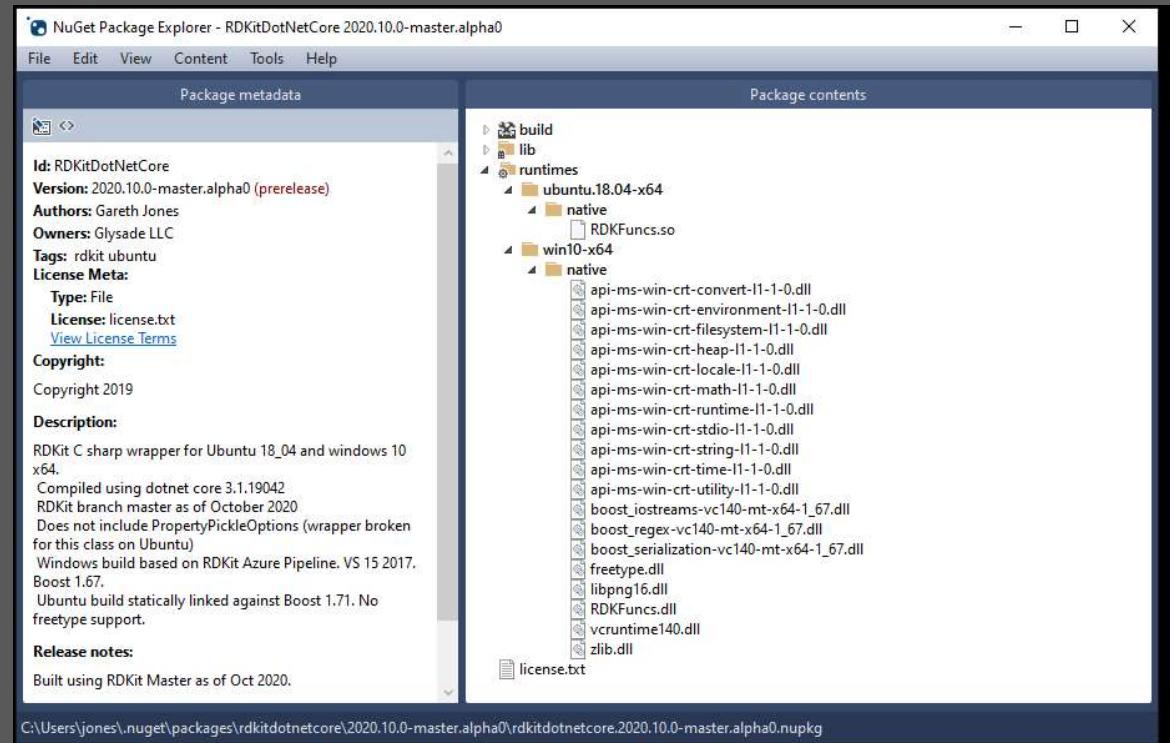


Lead Discovery ChemCharts



RDKit within ChemCharts

- Use RDKit C# SWIG bindings
- Single NUGET package for both Windows and Linux
- Extend C# Spotfire extension



Web Services

- Asynchronous services for functionality that cannot be provided within Spotfire

Lead Discovery ChemCharts

Column 1	MOLECULE_ID	Avg Log IC ₅₀	SMILES	
1	BD441129	9.69		
2	BD441181	9.69		
3	Neighbor Properties			
4	Add nearest neighbors information to a datatable with a structure column			
5	Data source for new task	<input checked="" type="radio"/> Data table <input type="radio"/> Previous task	BeaconsAR_Assay/654 - IC ₅₀ results for target WT 19A/J	
6	Select structure column	SMILES		
7	ID column(s)	MOLECULE_ID		
8	Data columns	Avg Log IC ₅₀		
9	Submit data from	<input type="radio"/> All data <input checked="" type="radio"/> Data after filters <input type="radio"/> Marked data		
10	Task name	Enter name for task		
11	<input type="button" value="Submit"/> <input type="button" value="Cancel"/>			
12	8	BD440069	9.24	
13	9	BD441208	9.23	

Services

Search

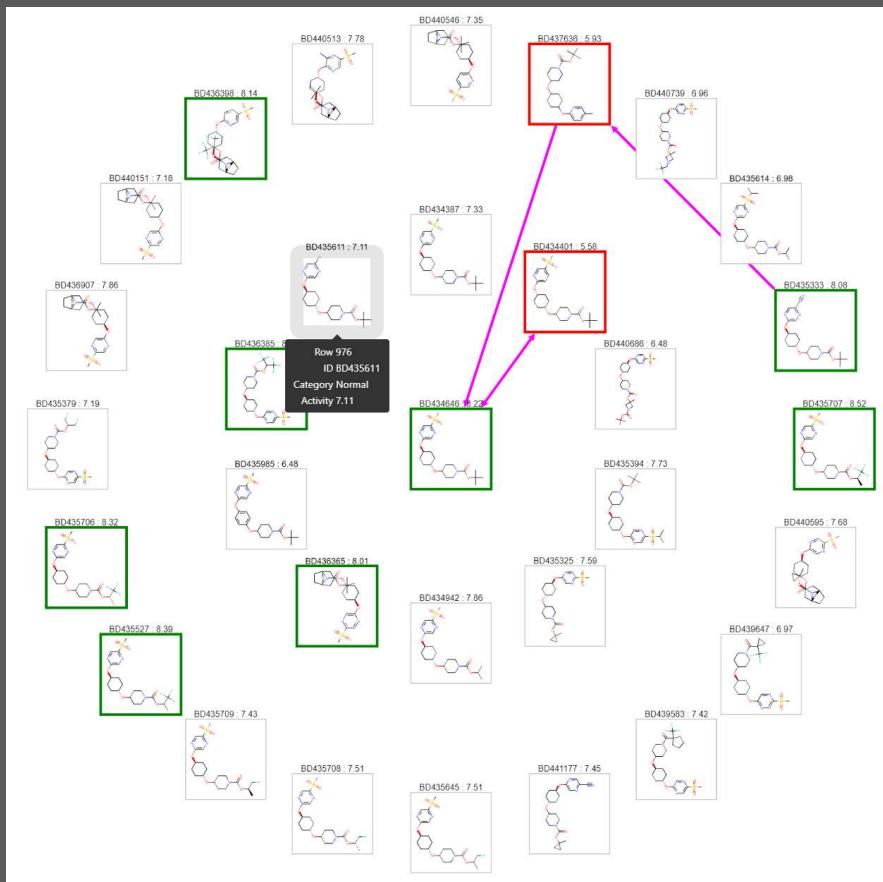
- bio
 - sequence
- chem
 - 2D
 - Compute Properties**
Computes a set of properties using Pipeline Pilot as a web service
 - Core Decomposition Analysis**
Perform core decomposition analysis on a structure column
 - MMP Database Search**
Finds MMP transformations and associated property predictions for a query structure
 - Neighbor Properties**
Add nearest neighbors information to a datatable with a structure column
 - R Group Analysis**
R Group Analysis Service
 - R Group Decomposition**
Performs R group decomposition on the structures in a data table column
 - RDKit Scaffold Network Analysis**
Perform core decomposition analysis on a structure column
 - RDKit Substructure Search**
Performs in-memory substructure search on specified resources
 - Similarity Search**
Searches databases for chemically similar structures
 - Substructure Search**
Searches databases for chemically similar structures
 - 3D
 - FastROCS Database Search**
Search shape databases using FastROCS
 - Omega2**
Generates single conformer for the 2D molecules in the table and adds them as a new column
 - Posit**
Runs POSIT on the molecules from the chosen column in a Spotfire table
 - ROCS Database Search**
Searches stored database using ROCS

Web Services Architecture

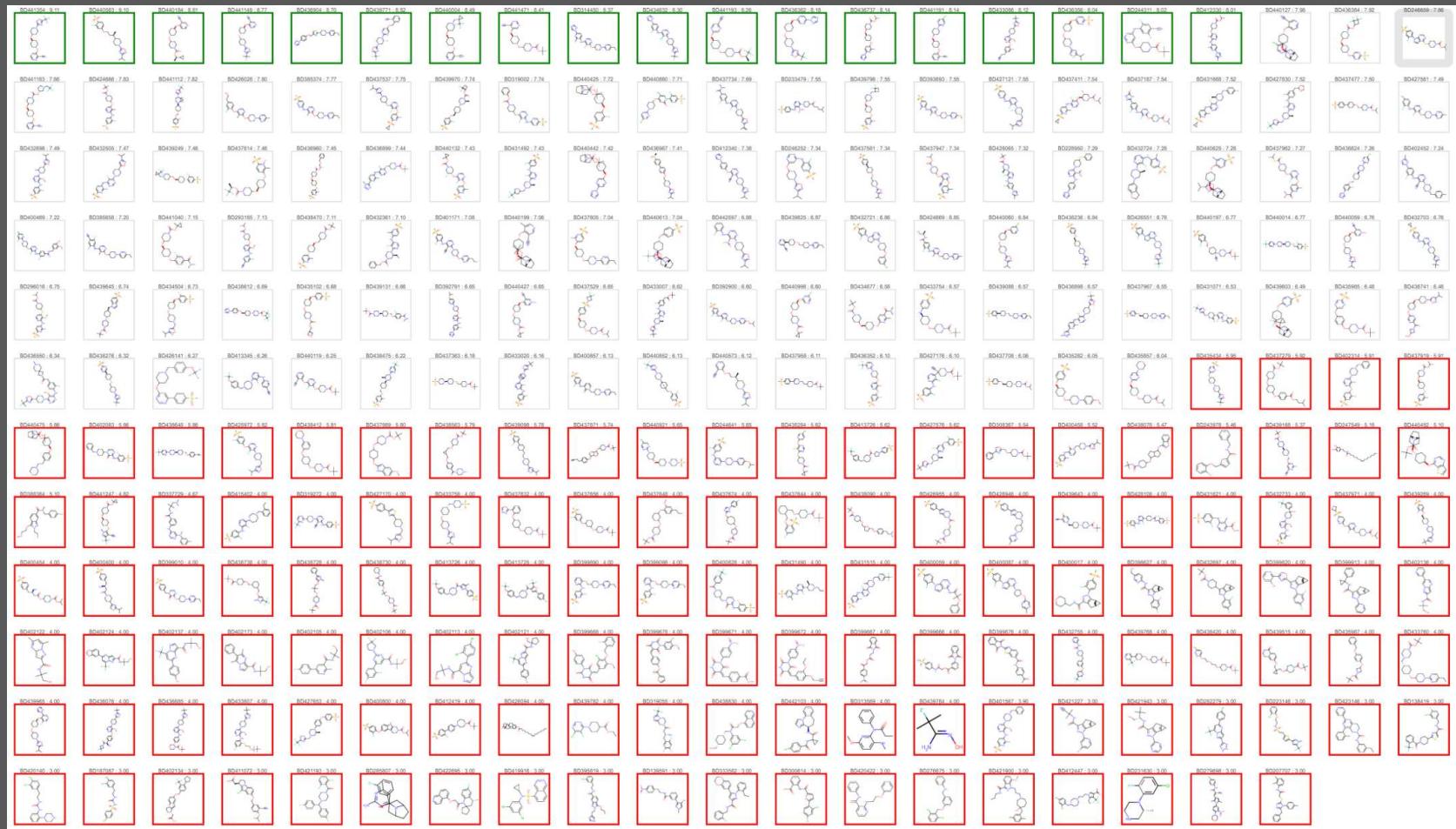
- Open architecture
- Single centralized server
- Distributed worker nodes
- Docker containers
- AWS hosting
- Extensible
- C#
- Python



Neighbors Properties Visualizer

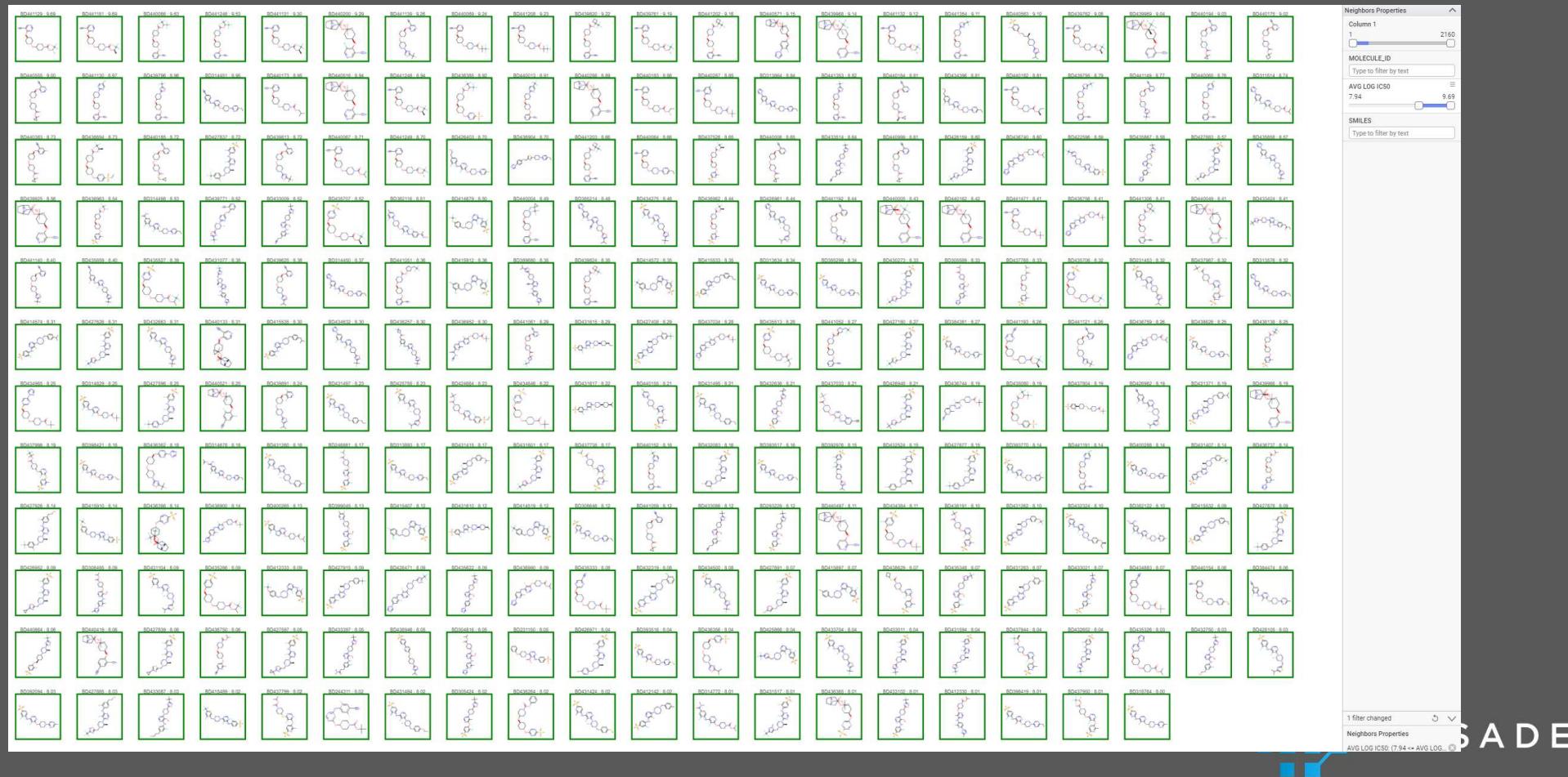


Landing Page



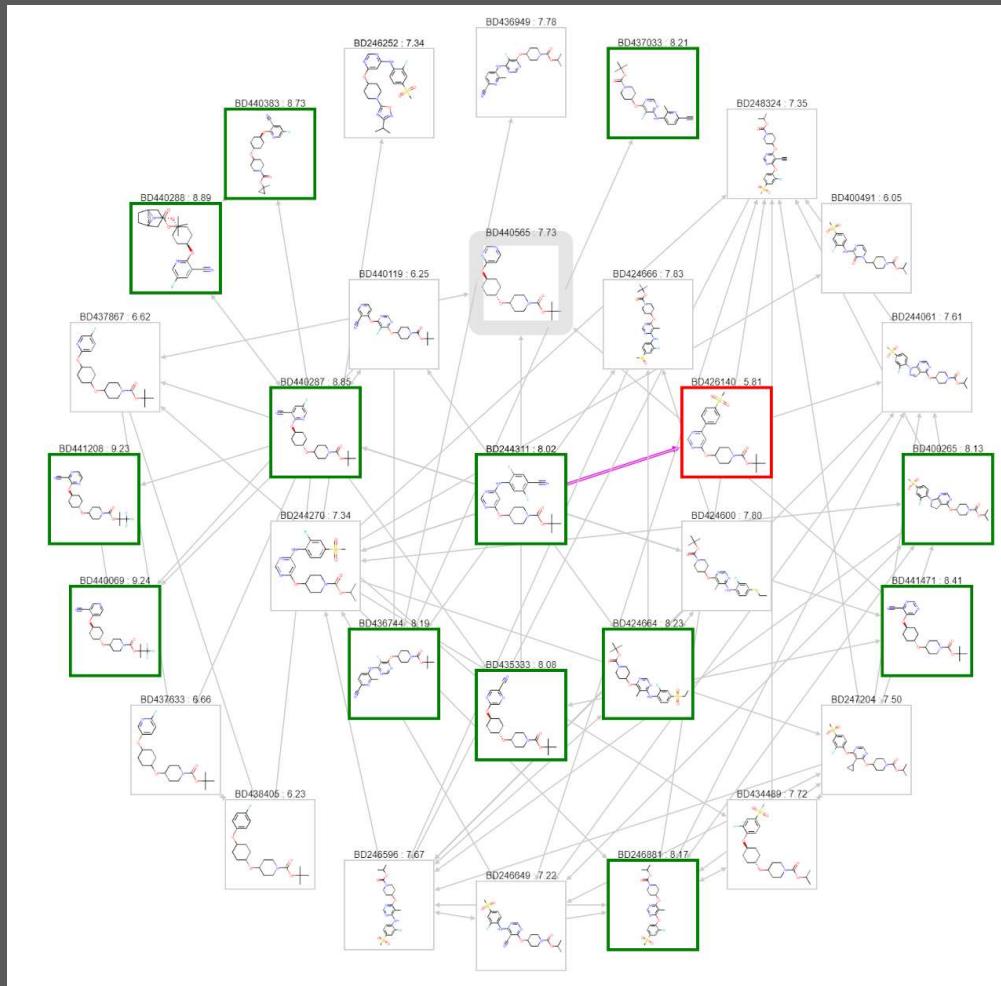
SADE

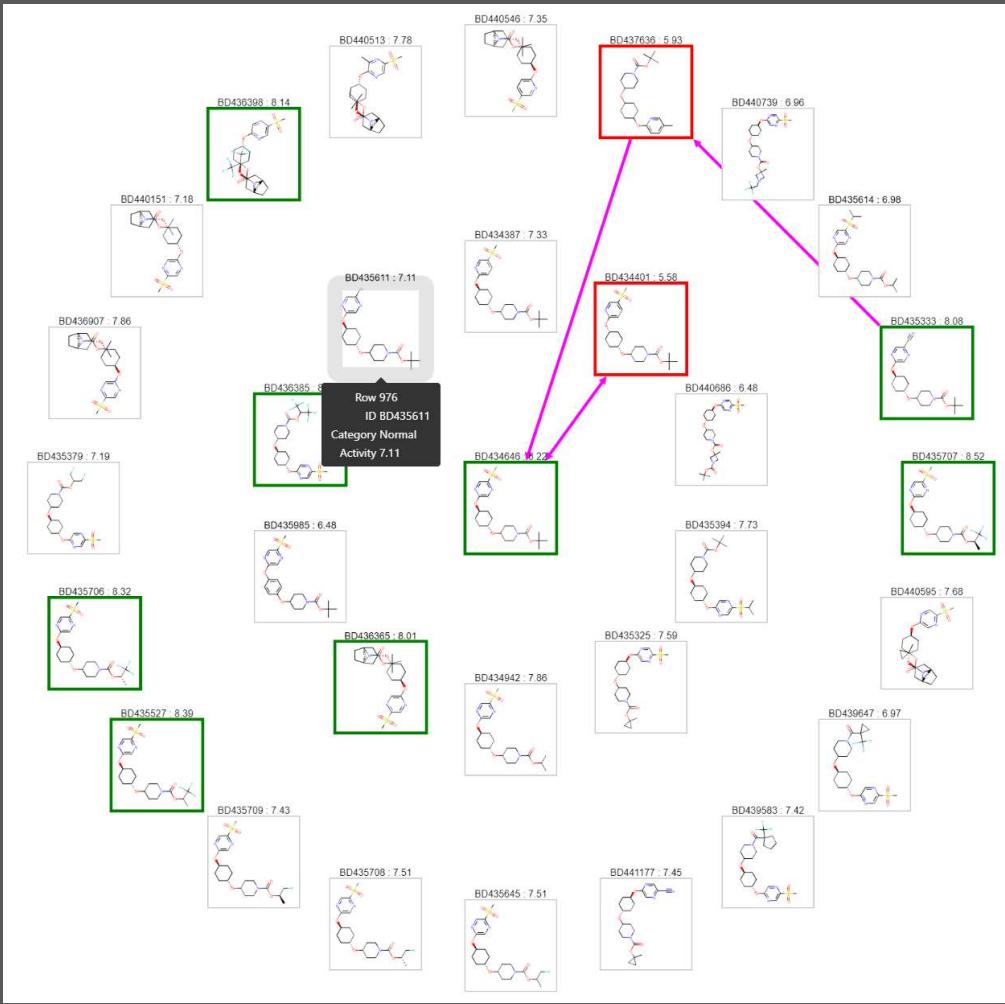
Filtered Landing Page



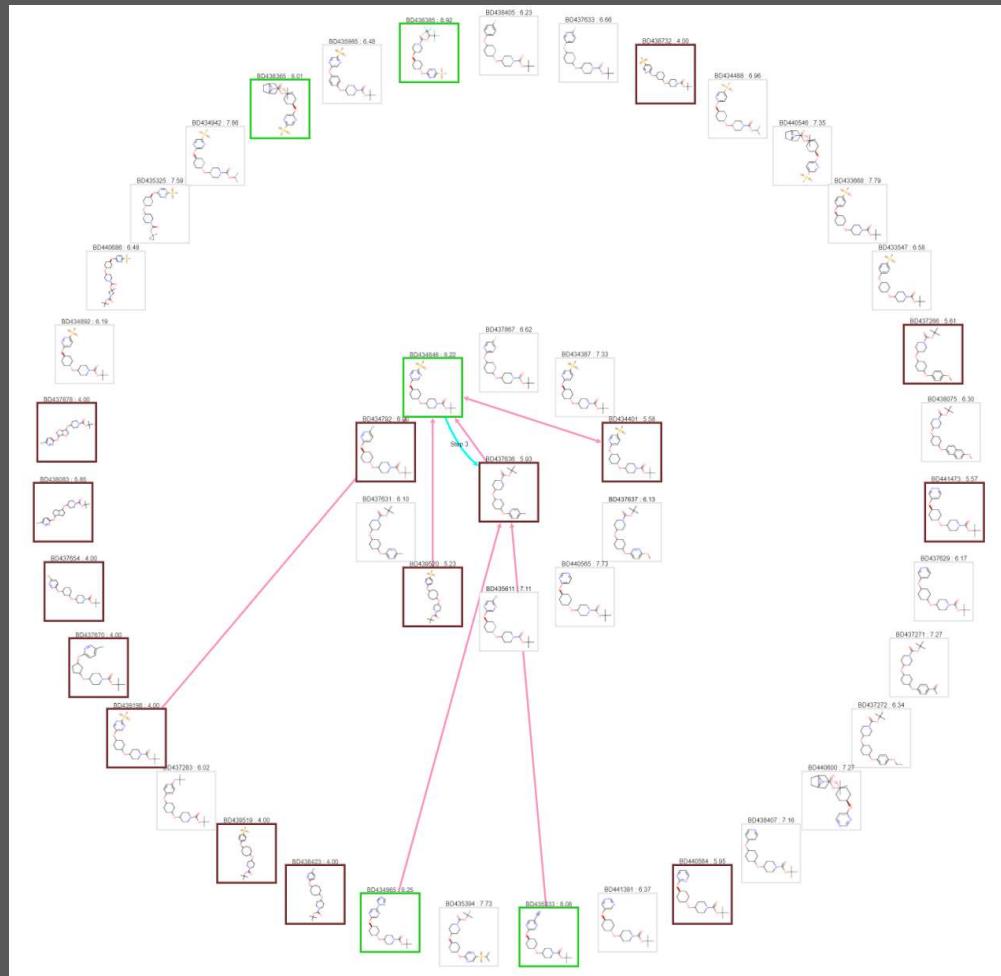
SADE

Neighbors view





Customize View



Neighbor properties

General

Data

Colors

Neighbor properties

Nearest neighbors column: NearestNeighborsInformation

Property column: AVG LOG IC50

ID column: MOLECULE_ID

Outer ring size: 45

Node borders: Use column colors for node borders

Edge labels: Use change delta for transition edge labels

Edge colors:

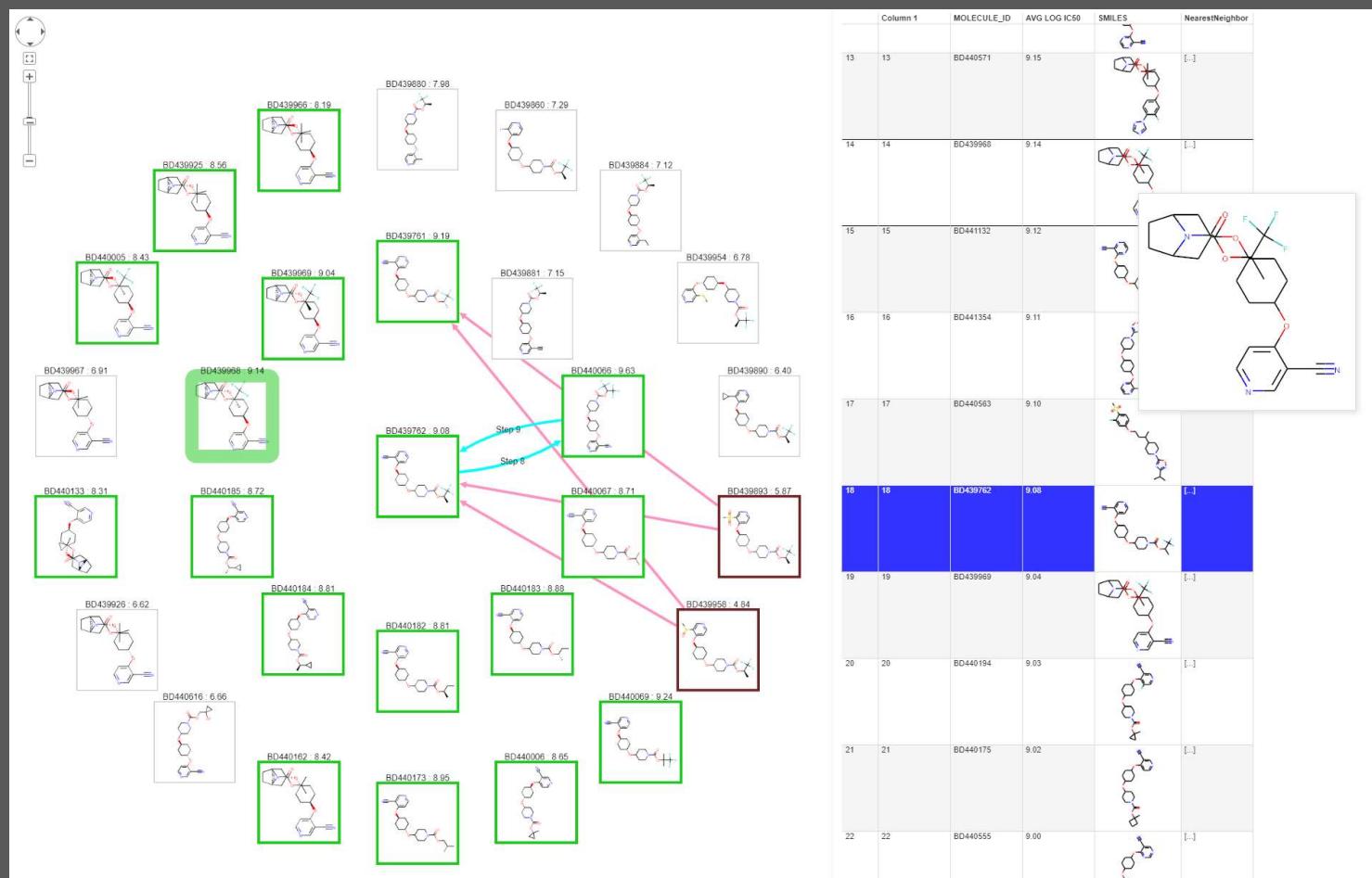
- Transition edge color: █
- Neighbors edge color: None User defined
- Path edge color: None User defined █

Categories:

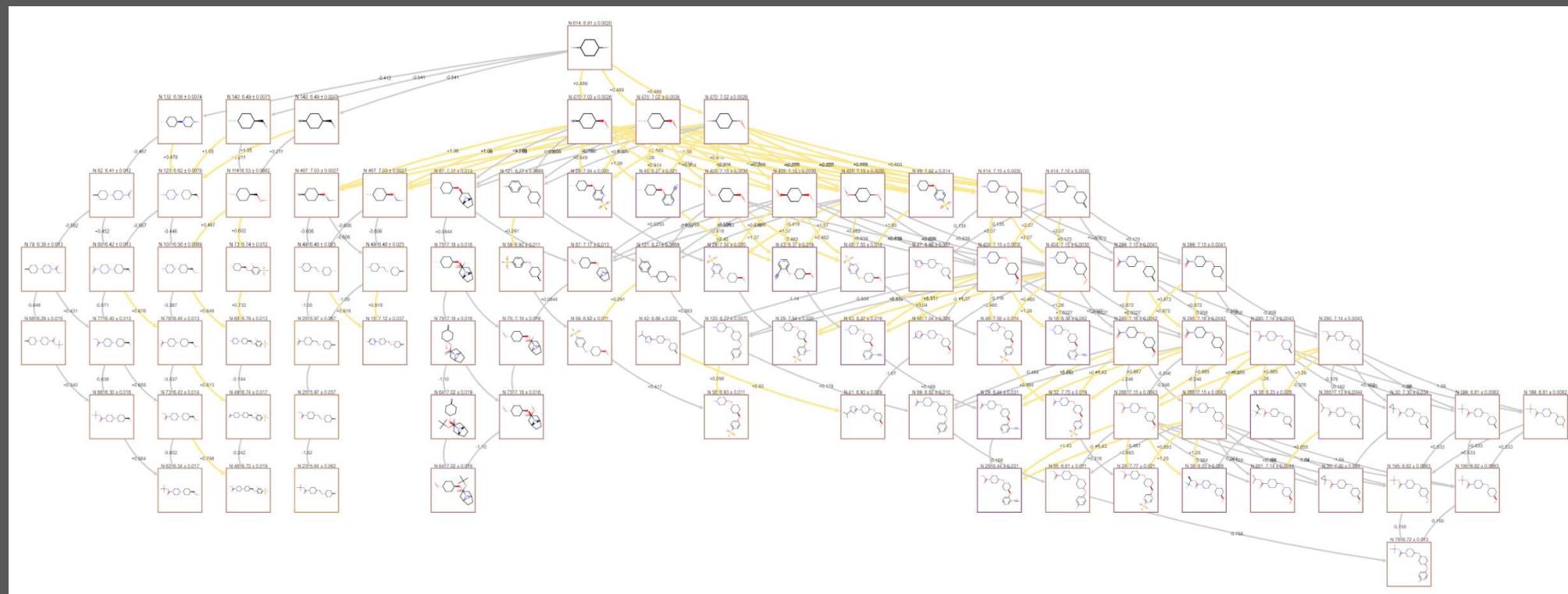
high	8	Value	█
low	6	Value	█



Data table integration

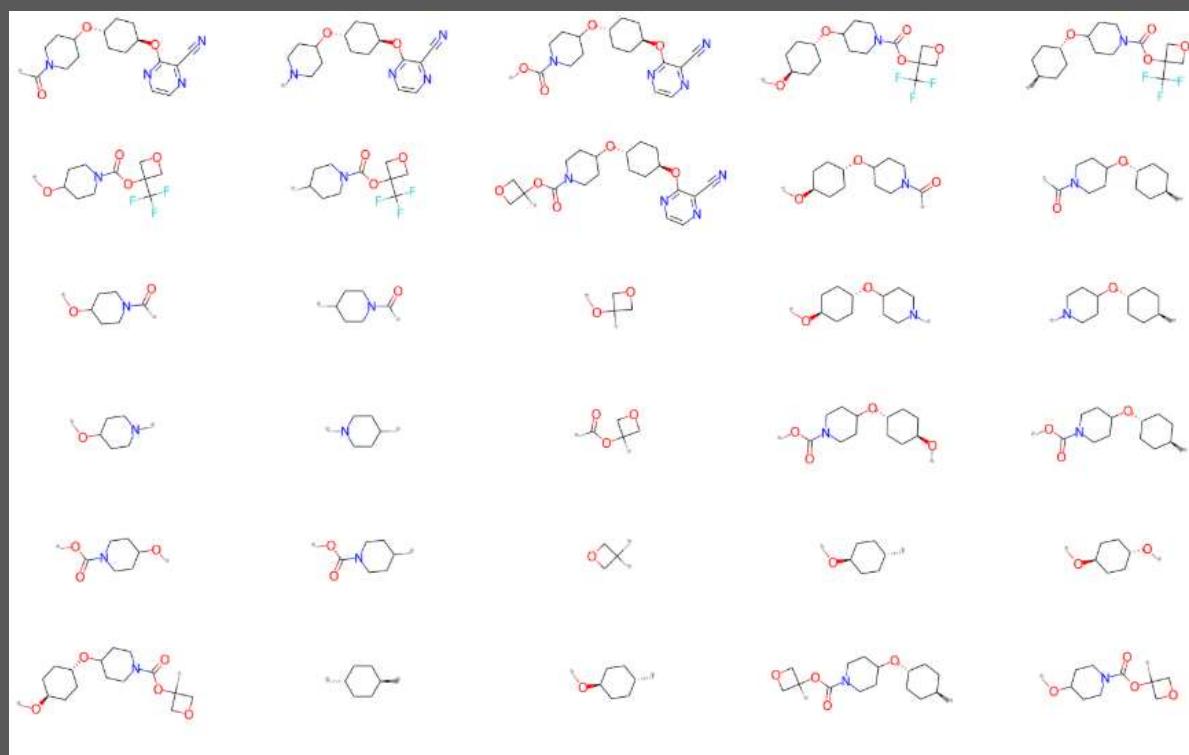
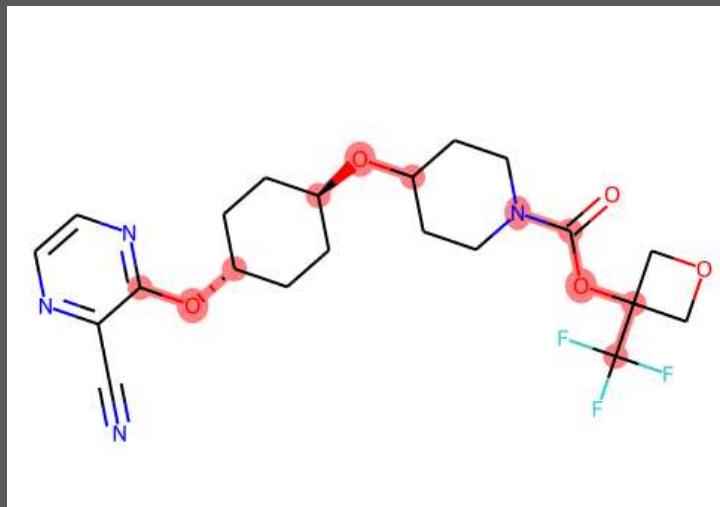


Structure Network Visualizer



Core Decomposition Service

- Decomposes molecules into a hierarchy of core fragments

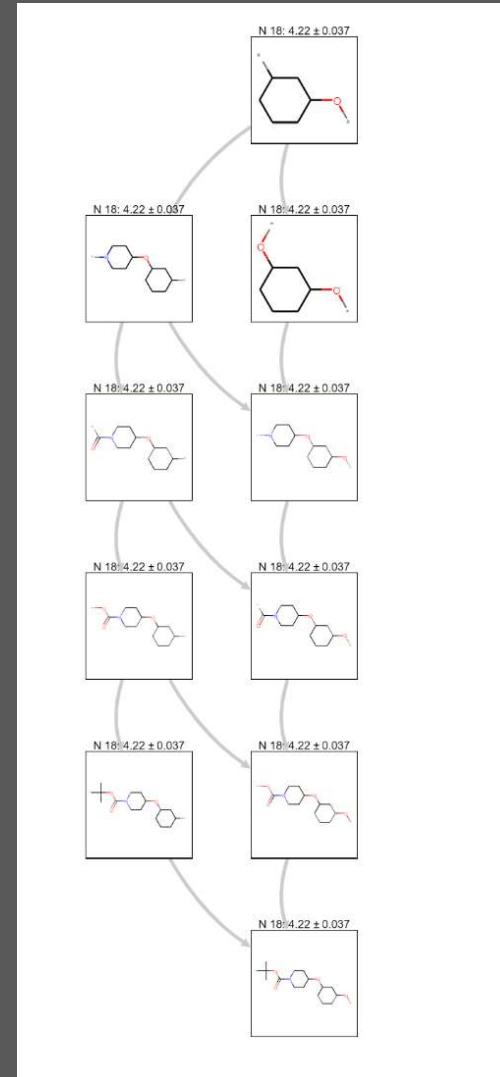


Wassermann, A. M.; Haebel, P.; Weskamp, N.; Bajorath, J. SAR Matrices: Automated Extraction of Information-Rich SAR Tables from Large Compound Data Sets. *J. Chem. Inf. Model.* 2012, 52 (7), 1769–1776. <https://doi.org/10.1021/ci300206e>.



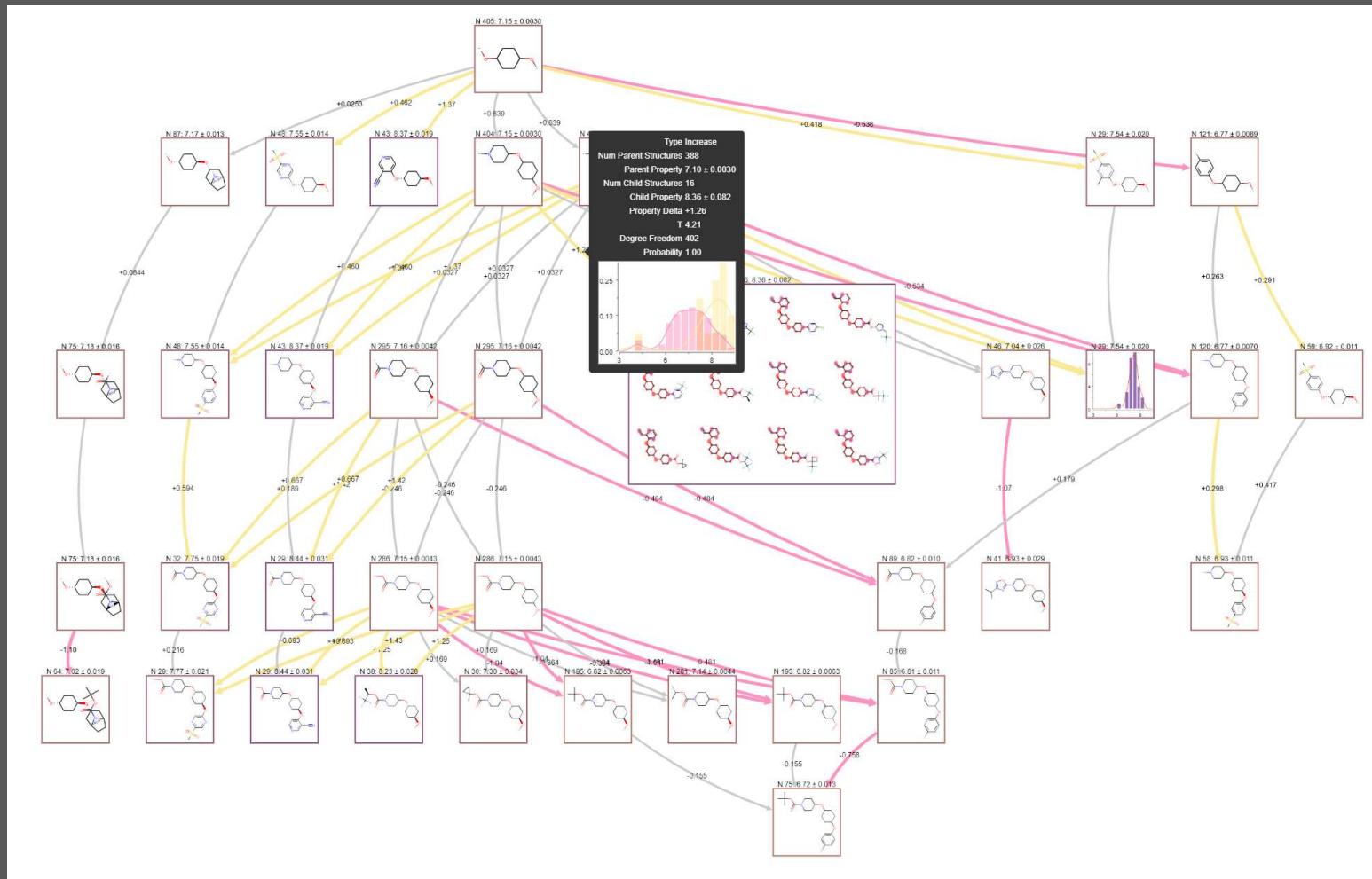
Core Decomposition Service

- Determine hierarchy: which cores are sub-cores
- Identify compounds which contain cores
- Retain the most populous 500 cores
- Output DAG of cores



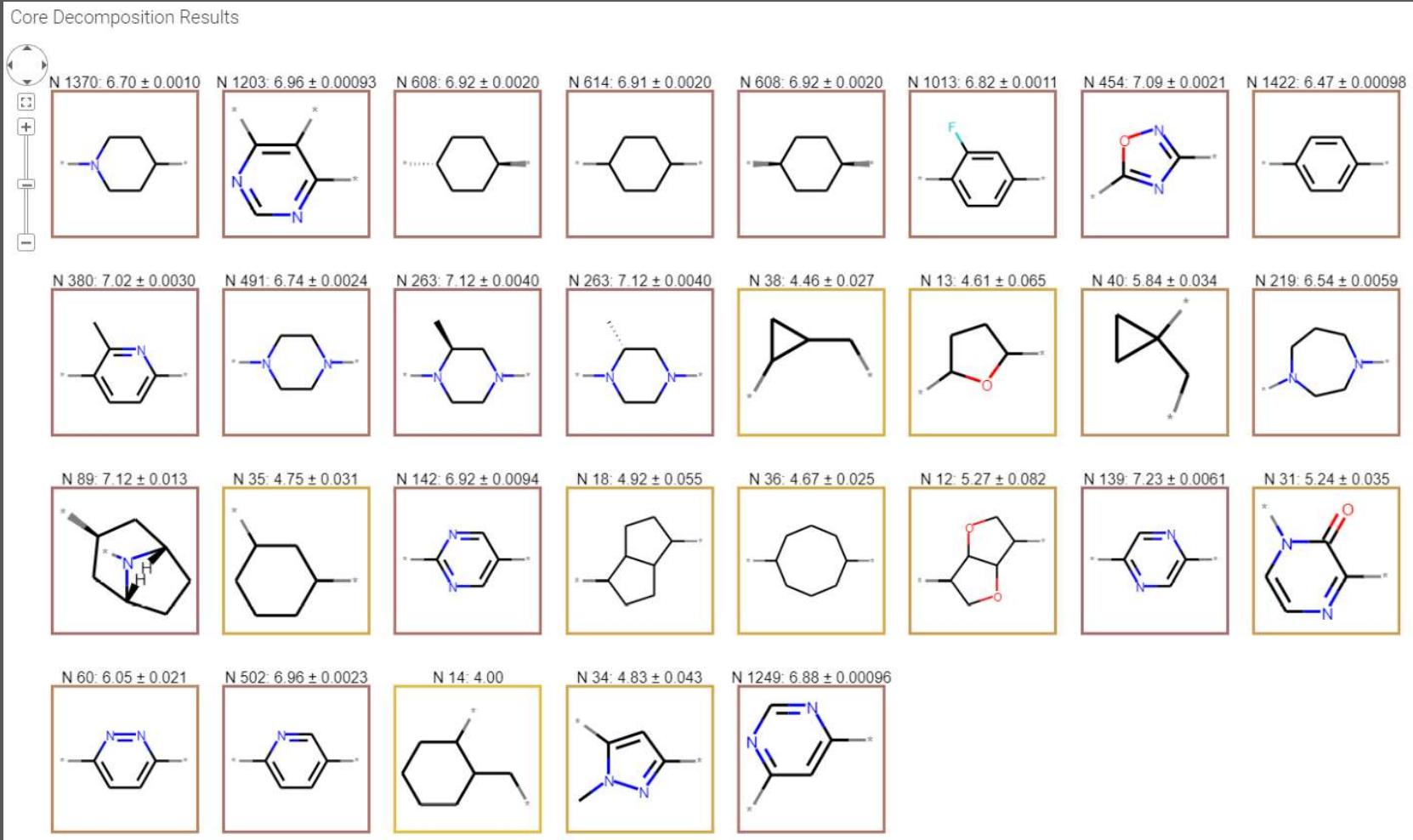
GLYSADE

Structure Network Visualizer



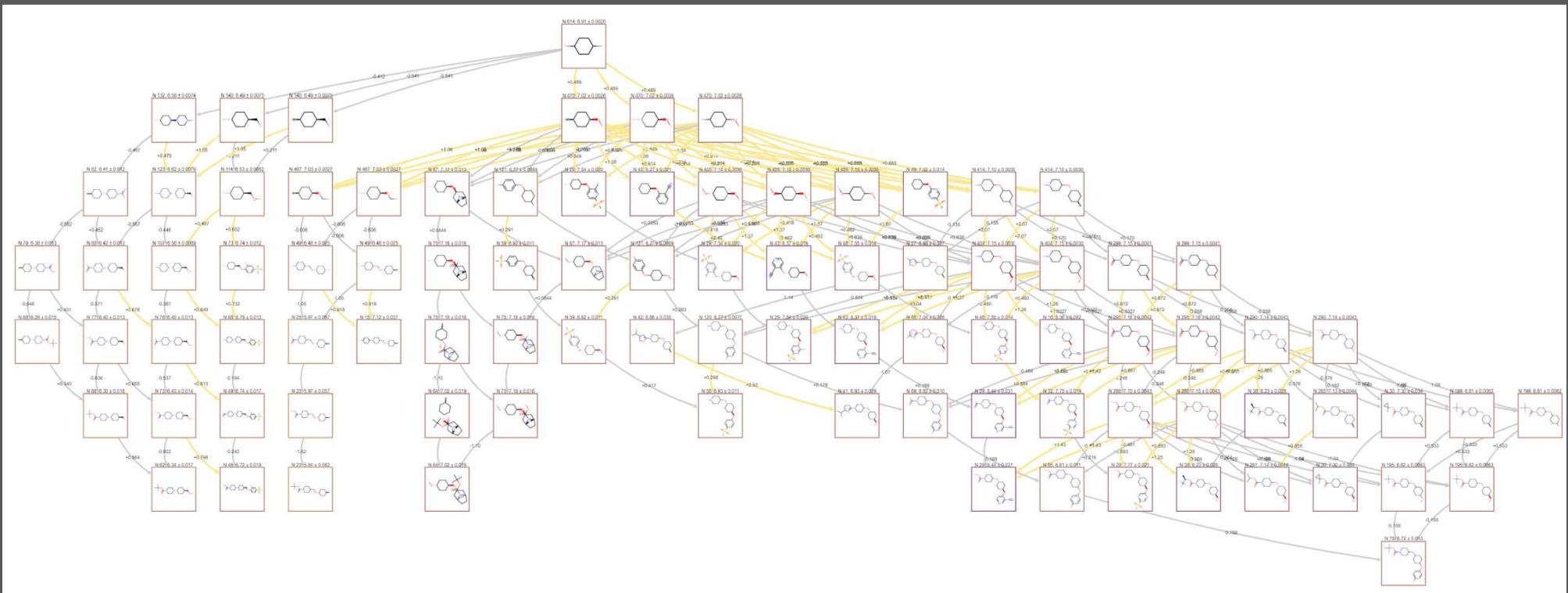
GLYSADE

Landing page of root nodes

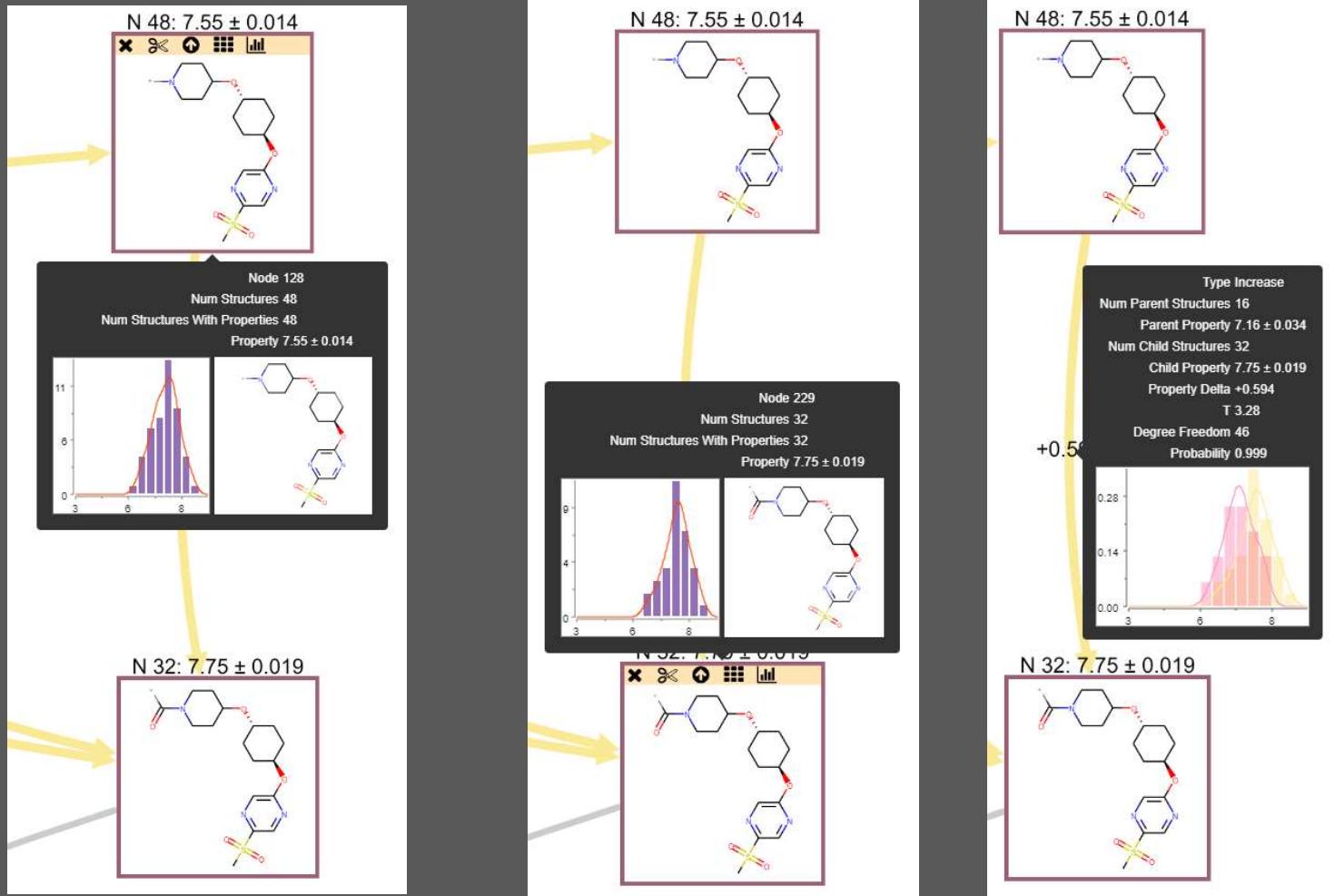


DAG View

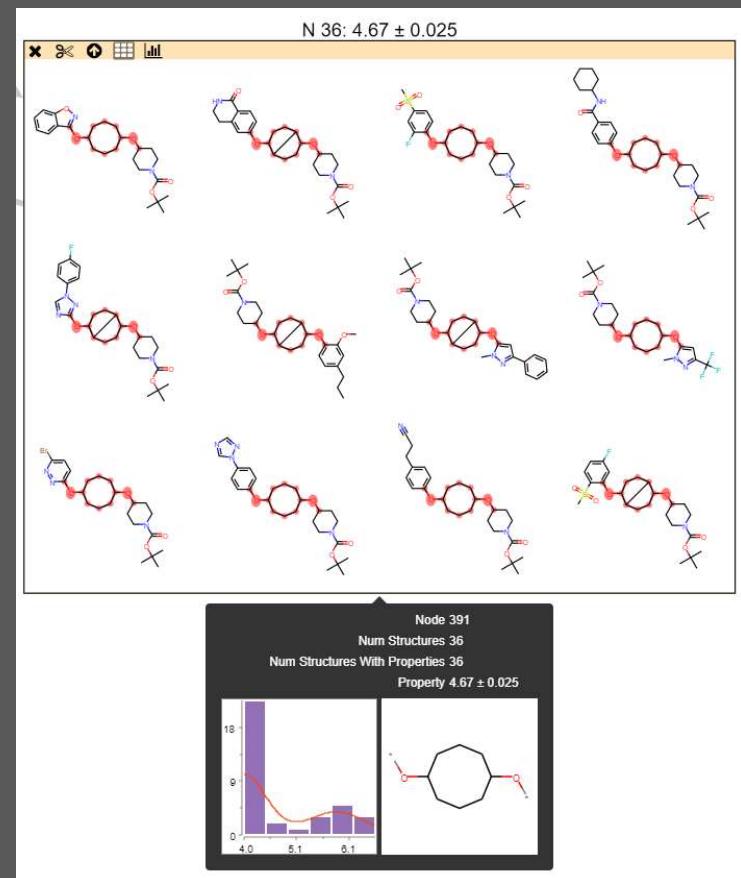
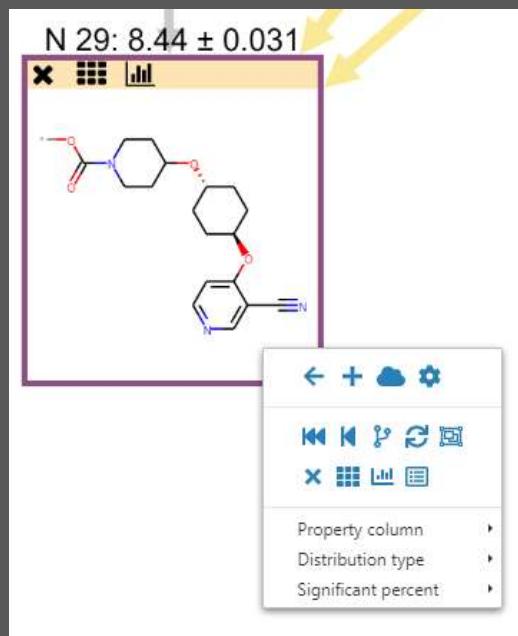
- Sugiyama layout
 - Aligned cores
 - Significant edges colored



Node and edge tooltips



Node based operations

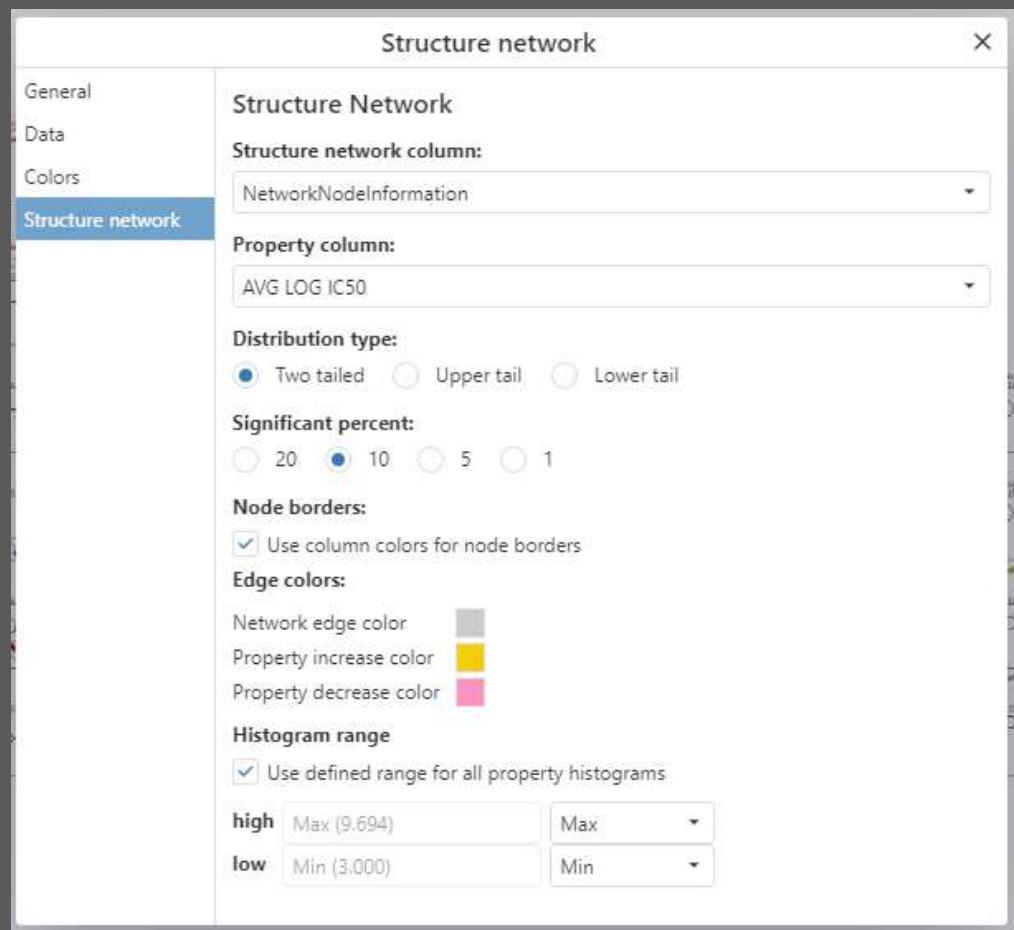
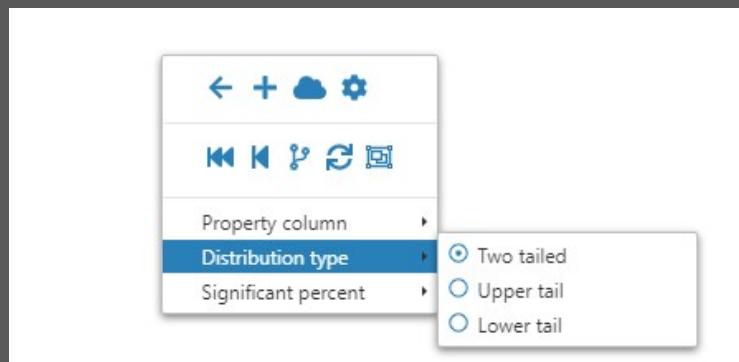


Layout operations



Customization

- Context menu
- Property pages

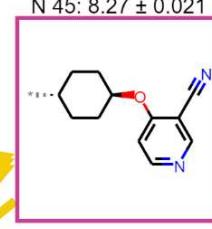


Data Table Integration

Lead Discovery ChemCharts

Column 1	MOLECULE_ID	AVG LOG IC50	SMILES	NetworkNodeInf
36	36	BD440184	8.81	[...]
37	37	BD434396	8.81	[...]
38	38	BD440182	8.81	[...]
39	39	BD439795	8.79	[...]
40	40	BD441149	8.77	[...]
41	41	BD440065	8.76	[...]
42	42	BD311514	8.74	[...]
43	43	BD440383	8.73	[...]

N 45: 8.27 ± 0.021



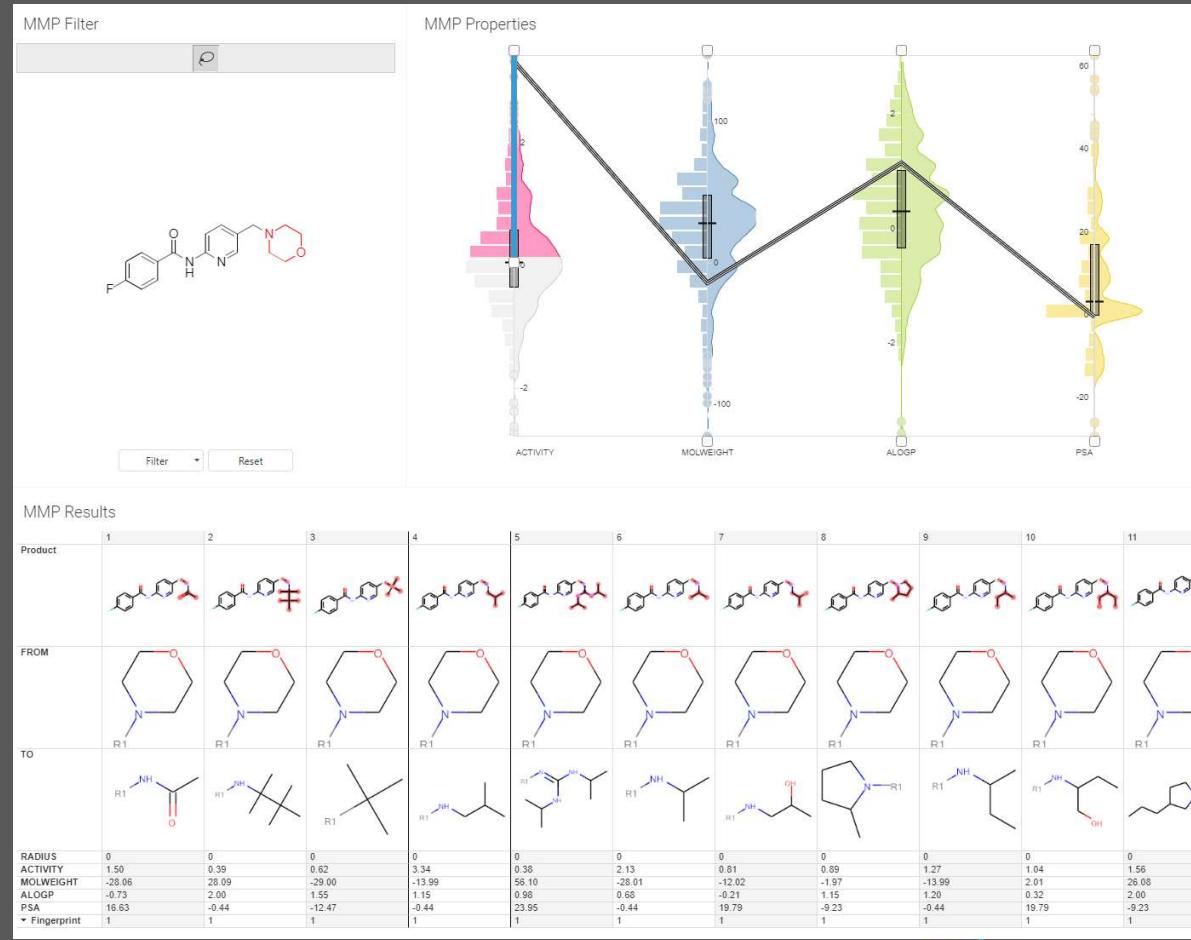
The chemical structure is a cyclohexane ring substituted with a 2-(4-cyano-2-methylphenyl)ethoxy group.

SMILES: C#Cc1ccc(OCC2CCCCC2)cc1

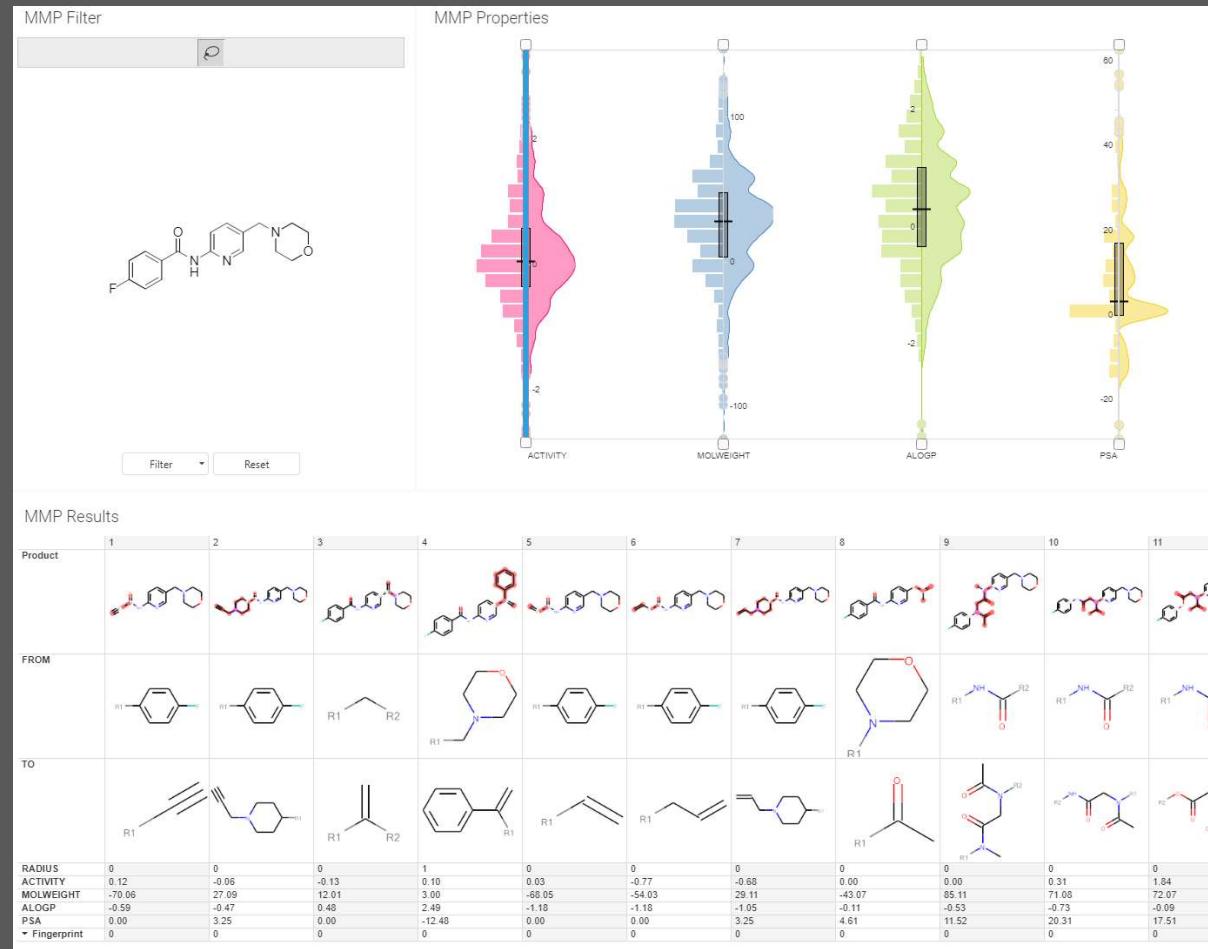


MMP Results View

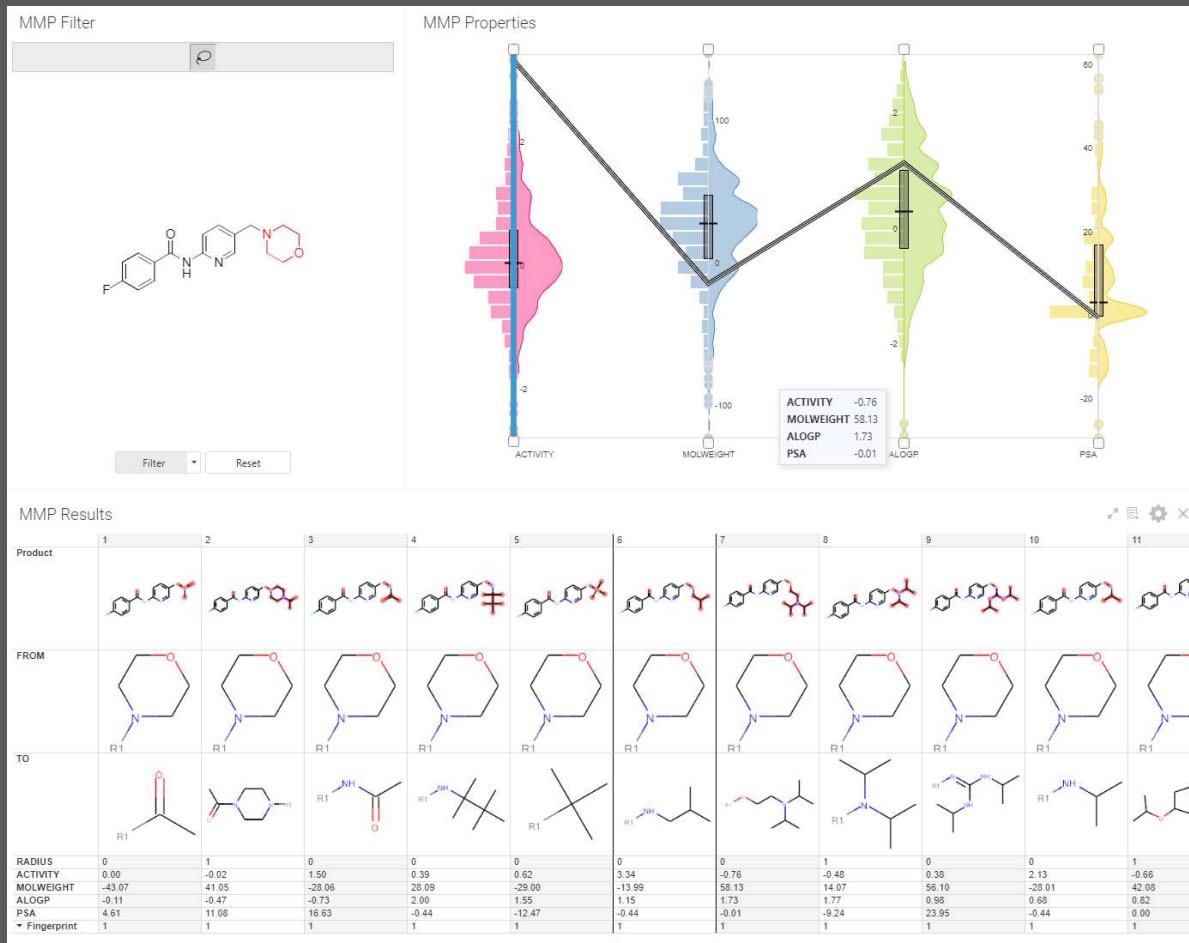
- MMPDB Web Service
 - Andrew Dalke



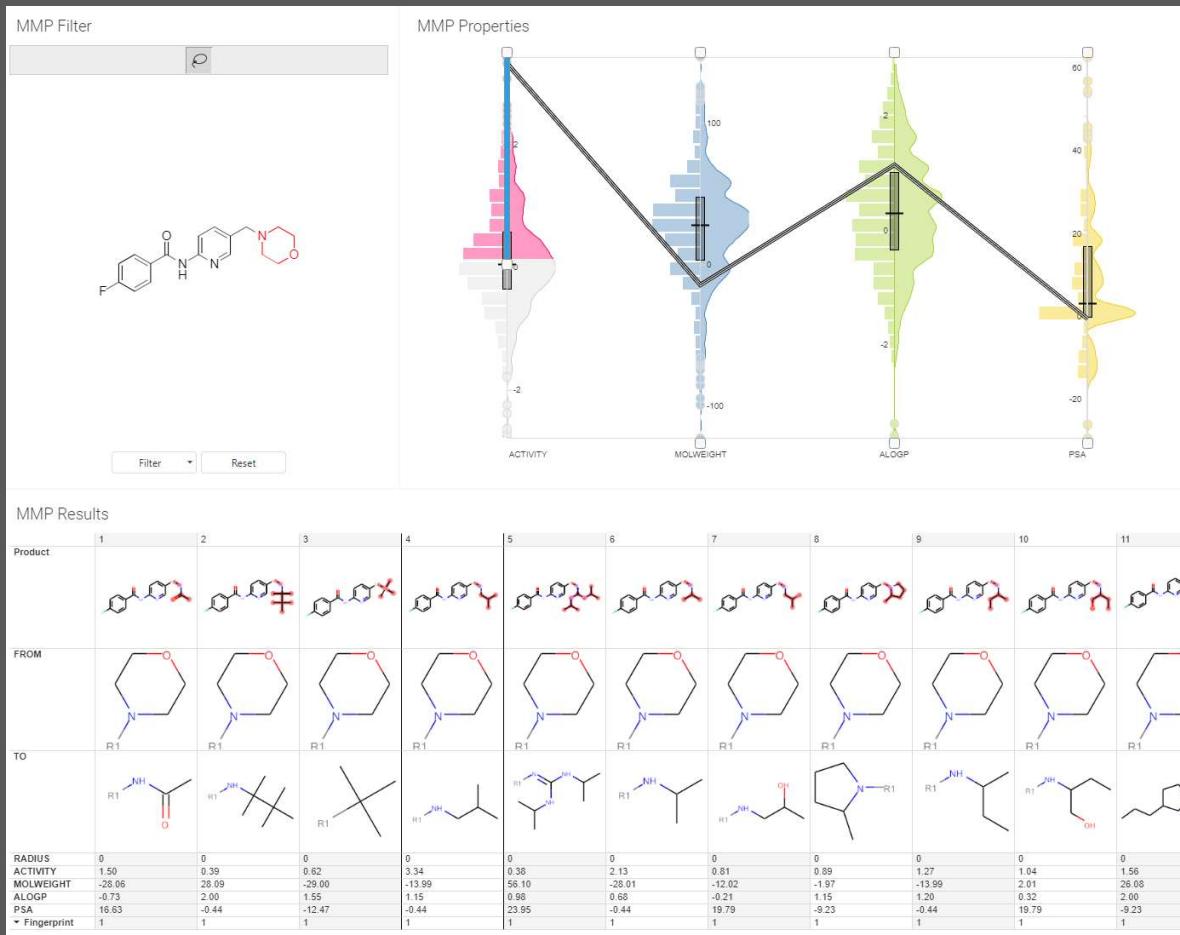
MMP Visual



Select substructure



Filter by property



Other RDKit projects

- Oracle cartridge/extensible index for substructure and similarity search
 - Java RMI
 - Funded by Beacon Discovery
 - <https://github.com/jones-gareth/OraRdkitCart>
- Development for Schrodinger Inc
 - Tautomer query search and fingerprinting in core RDkit



Acknowledgements

- GlySade
 - Matt Stahl
 - Paul Watson
- PerkinElmer
 - Dan Weaver
- Beacon Discovery
 - Carleton Sage
 - Dave Unett
- DART Neuroscience
 - Brock Luty
- Greg and RDKit community

